Learn to see

The vOICe Training Manual

**Last update November 24, 2018**

Updated versions of The vOICe Training Manual will be made available for download as a zipped Microsoft Word file with linked MP3 sound files, [manual.zip](https://www.seeingwithsound.com/manual.zip). Translations into [Russian](https://www.seeingwithsound.com/manual_ru/The_vOICe_Training_Manual_ru.htm), [Portuguese](https://www.seeingwithsound.com/manual/The_vOICe_Training_Manual_pt_br.htm) and [Chinese](https://www.seeingwithsound.com/manual/The_vOICe_%E5%9F%B9%E8%AE%AD%E6%89%8B%E5%86%8C_cn.htm) are also available.

|  |
| --- |
| **Want to make a difference? Set up a** External link[**vision training center**](https://voicevision-za.seeingwithsound.com) **for the blind in your country!** |

**Table of contents:**

[Learn to see 1](#_Toc530860091)

[The vOICe Training Manual 1](#_Toc530860092)

[1 Introduction 1](#_Toc530860093)

[2 Image to sound mapping principles 3](#_Toc530860094)

[2.1 Image examples with soundscapes 4](#_Toc530860095)

[3 Reaching and grasping 6](#_Toc530860096)

[4 Interpreting distance and size 10](#_Toc530860097)

[4.1 Changes in apparent size with distance 10](#_Toc530860098)

[4.2 Other distance and size clues: parallax and occlusion 11](#_Toc530860099)

[5 Visual perspective 13](#_Toc530860100)

[6 Visual landmarks 14](#_Toc530860101)

[7 Ground level hazards 14](#_Toc530860102)

[8 Training schedule 15](#_Toc530860103)

[9 Performance checklist 15](#_Toc530860104)

# Introduction

Sighted people often take vision for granted, because it seems so effortless. However, vision is in fact a highly complex skill that still largely defeats the best efforts in computer vision for object recognition in real-life situations, and neuroscience has shown that in sighted people a large part of the brain is directly or indirectly involved in processing input from the eyes. Moreover, vision is often inherently ambiguous, and it requires knowledge of the world, knowledge of context, and extensive visual experience to reliably disambiguate typical visual input from the environment. The vOICe sensory substitution technology now converts raw visual views into corresponding soundscapes while preserving a significant amount of visual information. Technically this allows you to see with sound, but your brain must first learn to decode the soundscapes in visually meaningful terms. This takes time and practice, because this skill is completely new to your brain, and the purpose of this manual is to provide you with a set of exercises and guidelines that help you acquire the basic skills needed to make sense of visual input encoded in sound. From there on, it is only through extensive and immersive use of The vOICe in real-life situations that you further master seeing with sound, to ultimately make it second nature. As such it is not unlike learning a foreign language, where you first learn a grammar and a vocabulary to form a solid basis, but subsequently make it increasingly effortless and subconscious through years of practical use. No one can learn a new language overnight, and similarly The vOICe is not a magic bullet for instant sight. Practice makes perfect, even if progress can seem slow at times, just as with learning a new language.

Technical details that are specific to a version of The vOICe running on a particular operating system (e.g. Microsoft Windows or Android, or look-alikes on iOS) or to particular hardware setups (e.g. camera glasses, a smartphone or augmented reality glasses) are omitted from this manual. The focus here is entirely on learning to interpret the soundscapes, and the same principles apply to all implementations of The vOICe in software and hardware. Please refer to the [seeingwithsound.com](http://www.seeingwithsound.com) website for implementation details and disclaimers, such as for [The vOICe for Windows](http://www.seeingwithsound.com/winvoice.htm) software and [camera glasses](http://www.seeingwithsound.com/camera_glasses.htm) for devices running Microsoft Windows, and [The vOICe for Android](http://www.seeingwithsound.com/android.htm) for smartphones and [augmented reality glasses](http://www.seeingwithsound.com/android-glasses.htm) running Android. Also, the focus of this manual is on vision skills relevant to ambulatory vision (moving around supported by vision based navigation, obstacle detection, as well as reaching for and grasping objects). The focus is not on any more specific uses such as reading newspaper headlines, door labels or reading graphs or weather maps. For sighted humans there is an almost unlimited number of use cases for vision, and for this manual we draw the line at the core vision skills, skills that are in fact not unique to humans and human culture but key to survival throughout the animal kingdom. After acquiring the core vision skills you should be better able to work independently on additional skills that may be relevant to your personal situation, your interests and your work.

If you are not totally blind you must blindfold yourself, or in the case of using camera glasses add clip-on sunglasses that are made completely opaque through black tape, such that you cannot make use of any residual eyesight in performing the exercises. Also, do not make use of echolocation in performing the exercises. Of course, once you have thoroughly mastered the exercises you can and probably should again make use of any supplemental clues from residual eyesight or echolocation, but just not during the exercises.

Concerning audio volume, do not annoy and tire yourself with loud soundscapes. Use the softest level that still works for you, because you learn best when you can maintain an active interest in subtle audio clues. You may develop a still higher sensitivity for sound that will make you prefer very modest sound levels, while using low sound levels minimizes nonlinear distortion as well as long-term risk of hearing damage. Completely mute The vOICe when that is best and safest for the situation at hand.

|  |
| --- |
| **The Need for Speed** |
| **A common error in learning to use The vOICe is to keep acting slowly, all the time consciously and carefully analyzing the view. At some point, you *must* skip conscious analysis to become fluent and learn to perform actions at speed. If safety permits, act fast. Sighted people do not think about vision when reaching for something: they just do it!** |

# Image to sound mapping principles

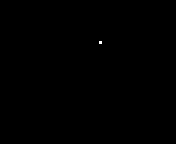
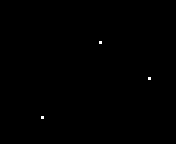
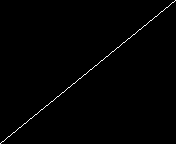
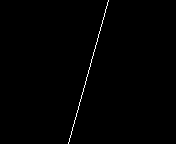
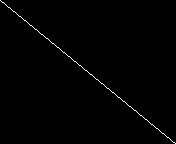
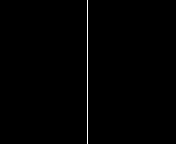
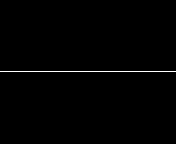
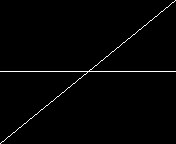
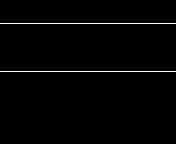
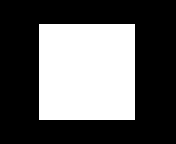
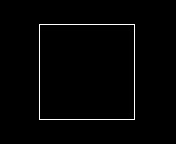
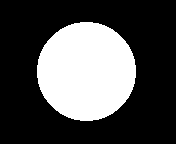
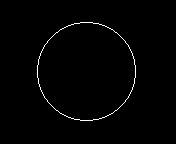
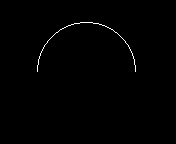
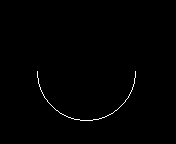
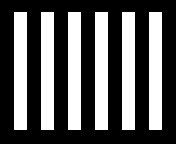
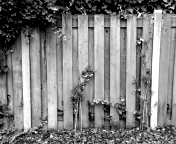
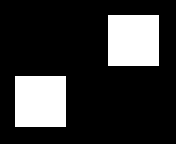
There are only *three simple rules* in The vOICe’s general image to sound mapping, each rule dealing with one fundamental aspect of vision: rule 1 concerns left and right, rule 2 concerns up and down, and rule 3 concerns dark and light. The actual rules are

1. *Left and Right.*  
     
   Views are sounded in a left to right scanning order, by default at a rate of one image snapshot per second. You hear the stereo sound correspondingly pan from left to right to make the scanning easier to track. Hearing some sound on your left or right thus means having a corresponding visual pattern on your left or right side, respectively.
2. *Up and Down.*  
     
   During every scan, pitch means elevation: the higher the pitch, the higher the position of the visual pattern. Consequently, if the pitch goes up or down, you have a rising or falling visual pattern, respectively.
3. *Dark and Light.*  
     
   Loudness means brightness: the louder the brighter. Consequently, silence means black, and a loud sound means white, and anything in between is a shade of gray.

In other words, *The vOICe scans the view from left to right, while associating height with pitch and brightness with loudness*. Another way of describing the mapping is that each view is scanned in thin vertical slices, starting with a vertical slice sounding on your left side and ending with a vertical slice sounding on your right side. At any moment, the generated sound depends on the visual content of the current vertical slice, with higher pitched tones for bright pixels at higher positions. The described mapping is universal and can represent any grayscale image. Images, such as photographic snapshots taken by a camera, are inherently two-dimensional and do not explicitly represent distance, but later on in the section “[Interpreting distance and size](#_Ref366322958)” we will discuss how to obtain and interpret distance clues using The vOICe.

## Image examples with soundscapes

The easiest and quickest way to better grasp the image to sound mapping principles is by considering a few simple image examples, starting with their visual description, analyzing what the visual view “should” sound like, and listening to some sample soundscapes:

* A pitch black background gives no sound at all. Black is silent.
* A bright dot gives a short beep, with pitch telling elevation; with more than one bright dot in the view you hear multiple short beeps.  
    
  [](file:///T:\HTML\vOICe\manual\dot.mp3) [](file:///T:\HTML\vOICe\manual\3dots.mp3)
* A rising (falling) bright line gives a rising (falling) tone; the steeper the line the faster the pitch of the tone changes.  
    
  [](file:///T:\HTML\vOICe\manual\lineup.mp3) [](file:///T:\HTML\vOICe\manual\linestup.mp3) [](file:///T:\HTML\vOICe\manual\linedown.mp3)
* A vertical bright line gives a click.  
    
  [](file:///T:\HTML\vOICe\manual\linevert.mp3)
* A horizontal bright line gives a constant tone. A horizontal bright line and a rising bright line give two simultaneous tones, one constant and one rising.  
    
  [](file:///T:\HTML\vOICe\manual\lineflat.mp3) [](file:///T:\HTML\vOICe\manual\lineflup.mp3)
* Two bright lines positioned above each other give two simultaneous tones. Three bright lines positioned above each other give three simultaneous tones.  
    
  [](file:///T:\HTML\vOICe\manual\lineflat2.mp3) [](file:///T:\HTML\vOICe\manual\lineflat3.mp3)
* An upright bright filled square or rectangle gives a noise burst that starts and stops suddenly. Overall pitch and pitch range tell elevation and height. An upright bright open square or rectangle gives the sound of a vertical bright line for the left edge followed by the sound of two horizontal bright lines for top and bottom edge and finally the sound of a vertical bright line for the right edge. A [video](file:///T:\HTML\vOICe\manual\squareani.wmv) with five soundscapes illustrates how construction of a filled square can be thought of as first drawing the bottom edge, then the right edge, top edge and left edge, and finally filling the open square.  
    
  [](file:///T:\HTML\vOICe\manual\square1.mp3) [](file:///T:\HTML\vOICe\manual\square0.mp3)
* A bright filled circle sounds like a noise burst that starts and stops gradually. A bright open circle sounds like two simultaneous tones, one tone for the top half circle going up and down, and one tone for the bottom half circle going down and up.  
    
  [](file:///T:\HTML\vOICe\manual\circle1.mp3) [](file:///T:\HTML\vOICe\manual\circle0.mp3) [](file:///T:\HTML\vOICe\manual\circle0tophalf.mp3) [](file:///T:\HTML\vOICe\manual\circle0bottomhalf.mp3)
* A set of equally-spaced bright vertical bars gives a constant rhythm of noise bursts. A pattern like this can be found in for instance a fence or in the pillars of the White House in Washington DC.  
    
  [](file:///T:\HTML\vOICe\manual\pillars.mp3) [](file:///T:\HTML\vOICe\manual\fence.mp3) [](file:///T:\HTML\vOICe\manual\whitehouse.mp3)
* Two upright bright filled squares or rectangles sound as two noise bursts.  
    
  [](file:///T:\HTML\vOICe\manual\squares.mp3)

The images embedded in the above examples link to MP3 sound files with corresponding soundscapes, such that CTRL-clicking an image will play its soundscape (or in the web version of this document you can just click the image links). Note that in Windows Media Player you can turn repeat on to have the soundscape loop just like The vOICe does when the image content doesn’t change. This will help you hear out details that might otherwise evade you when hearing a soundscape only once.

With the above in mind you can start exploring what real-life objects look (sound) like. For example, put your white cane on a dark surface, and notice how it gives a rising or falling tone depending on its orientation. A few shiny coins on a dark surface give a few corresponding beeps corresponding to their configuration. Try the rhythm made by the spines of books on a bookshelf, or the smoother rhythm made by the folds of closed curtains. Notice how a window as seen from inside a room typically sounds like a bright rectangle because outdoor ambient light is often brighter. In fact many man-made structures in an urban environment appear rectangular when viewed head-on: not only windows, but also doors and buildings.

When you hear a soundscape and verify through touch or other means what it represents, try to understand why it sounds like it does. If there is a tone, perhaps there is an edge. If there is a rhythm, perhaps there is a vertical grid. If there is a smooth noise, perhaps there is a smooth surface. Conscious, rational analysis will help you with the interpretation of views that you do not readily recognize, while paying attention to detail will sharpen your listening skills. Always be curious and eager to learn!

# Reaching and grasping

One of the most important practical skills to master is the ability to reach out and grasp an object. Using The vOICe, you can now learn do this visually, without a need for sweeping or groping: you “see” the object in the soundscape, and then get it in one directed movement. In order to do so, you must first master camera-hand coordination, just like the sighted have eye-hand coordination. This means that you learn what point in a soundscape corresponds to what viewing direction, and acquire the motor skills to accurately reach in that particular direction, in the end even without thinking about it. The description in this section assumes that you are using a head-mounted camera with the camera pointing in the direction of your nose, but the same procedures can, with appropriate changes in the description, be applied to the use of a hand-held camera – such as with a smartphone. However, a head-mounted camera, preferably in the form of camera glasses, provides the most intuitive and consistent viewing experience and is highly recommended for serious use.

Prerequisites for practicing reaching and grasping are:

* A dark tabletop, such as a tabletop made of dark wood, or a table covered by dark cloth. Non-reflective black, as with black felt, is preferred for best visual contrast.
* A few bright objects that you can cast on the table and reach for. Bright yellow or white DUPLO bricks make an excellent choice, but you could also cut palm-sized objects out of white Styrofoam packaging material, as long as the objects give good visual contrast with the dark tabletop and can be cast like dice without rolling off the table. In some of the details of the description we assume that you are using DUPLO bricks, which have a rectangular 3D shape with eight bumps on the top side.

Be seated at the dark tabletop that is emptied of anything visually distracting. Make sure that lighting is adequate for “seeing” the selected bright objects when placed on the table top, and check that the table surface appears void (with silence or soft noise in the soundscapes) when there are no objects placed on it. The dark surface serves as a non-distracting visual background, letting you focus your attention on the objects of interest and their positions in the camera view.

Now you drop one of the bright objects on the table such that it bounces around a bit without dropping off the table, landing at a somewhat arbitrary (random) position on the table, and your task is to grab the object without sweeping your hand over the table. So you need to visually locate the object and then reach out with your arm and grab the object with your hand. The most reliable way to do this is to first center the object in your (camera) view, and only then reach out.

In order to do so, you first have to get the object in the camera view if it is not already showing, by looking around. A small bright object sounds much like a beep. Once you notice the beep, you must center it in your camera view, both vertically and horizontally. Then the object is located right ahead, in the direction where your nose is pointing. An object is at the center of a soundscape view when it sounds halfway through the left-to-right scan (such that it sounds straight ahead), and at medium pitch. This centering alignment will now be described in more detail.

[](file:///T:\HTML\vOICe\manual\duplo5.mp3)

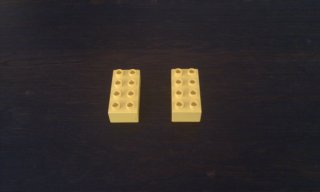
For vertical alignment, you need to tilt your head up and down until the object beep is at medium pitch (neither high nor low). Next, while maintaining this pitch, you turn your head left and right until the beep sounds half a second after the start of each soundscape, that is, in the horizontal middle of the default one-second duration of each soundscape scan. The direction from which the object sound seems to come will then also be straight ahead and not to the left or right. Then you can reach out and grab the object, imagining that it is in the direction where your nose is pointing. (Again, in case you are working with a hand-held camera instead of camera glasses, imagine the viewing direction of the camera from how you are holding it.)

In early practice, it is quite normal to be a bit off the mark, but already after a few hours of practice you should be able to grab the object spot on most of the time. Practice this exercise until it becomes a fluent action that you can repeat with ease. You can at the same time try to visualize the object that you are looking for in order to emphasize the visual nature of what you are accomplishing. Try to avoid falling into the old habit of sweeping your hand over the table to locate the object (although a quick correction of a few centimeters is OK). The goal is really to grasp for the object spot on, and this is entirely doable through practice. Learning camera-hand coordination can be fast because you can cast and grab for the DUPLO brick every five seconds or so, giving you some two hundred trials in just 15 minutes.

Moreover, you will notice that it is fairly easy to tell the orientation of a DUPLO brick from how fast its pitch rises or falls even though this takes only a split second. It can be much harder to tell if the brick landed upside down or not (or on its side), but the eight bumps on top of the brick as well as the pattern of diagonal ridges on the bottom both give a characteristic sound texture that contributes to the realism of “seeing”.

Once you have mastered the above, you can relax the condition of first centering the object in your view, and directly grasp for any off-center object, guided by pitch and the direction from where you hear the object sound. This is slightly harder, but also more efficient, because you can fetch the object from the first soundscape in which it appears, so in about a second. Reaching and grasping without first centering is therefore the preferred way of working, but you should only switch over to this after first mastering the centering of an object in your view.

The next stage is that you extend the single object grasping exercise by casting two or three DUPLO bricks onto the table, and grasp them all without first centering them one by one in your view. Thus, you can very efficiently grasp each of multiple objects from just a single visual sound view! Here too it is much like mastering a foreign language, where you first master conscious application of strict rules of grammar, but can later "forget" about the conscious application of these rules once it all becomes automatic and fluent, because conscious application would from then on only slow you down. In fact, at some point you *must* skip conscious analysis to become fluent. By analogy, no human can ride a bicycle by thinking about when to steer left or right: one would simply fall over. Sensorimotor skills must become largely subconscious and automatic to serve higher-level purposes that do require your conscious attention.

[](file:///T:\HTML\vOICe\manual\duplo_flash.mp3)

Once you can perform the tabletop grasping exercise with reasonable ease, you can start training for more mobile situations – but always in a safe (home) environment. Starting from a position at several meters distance from the table, center the object in your view, and walk towards it while keeping the object centered in your view. Note that you will need to bend over to prevent the object from vanishing from your view. Finally grasp it when you are close enough (with the object filling a noticeable slice of your view). In order to strive for maximum fluency and stimulate your eagerness to excel, think of this exercise as if you were a predator tracking and going after its prey. Initially the object will sound as a weak beep because it is still distant, but its appearance becomes more pronounced as you move closer and the object appears “bigger”, as demonstrated also by a [video](file:///T:\HTML\vOICe\manual\duploani.wmv) with five soundscapes.

[](file:///T:\HTML\vOICe\manual\duplo1.mp3) [](file:///T:\HTML\vOICe\manual\duplo2.mp3) [](file:///T:\HTML\vOICe\manual\duplo3.mp3) [](file:///T:\HTML\vOICe\manual\duplo4.mp3) [](file:///T:\HTML\vOICe\manual\duplo5.mp3)

To make the exercise more similar to typical daily living uses, you can place a white coffee mug or dinner plate instead of a DUPLO brick. The single object grasping exercise may sound elementary, and it is, but it is very important to first get the basics right, and it will also build confidence that good use can be made of certain aspects of the often overwhelmingly complex visual sounds. It is also part of many more complex behaviors and scenarios in mobile situations that can be mastered only after mastering this particular grasping skill. Reaching for a door handle is another immediate practical application of this skill. Directly grasping an object without groping or sweeping your hand is something that you could not have done without a form of sight.

*Initially practice the single object grasping exercise for half an hour daily, for at least two weeks.* This will give you a decent camera-hand coordination to start with and apply in daily life. In the longer run, you will still get much better at it, but there is a minimum skill level that you must reach in order to benefit from The vOICe.

Once you have thoroughly mastered this exercise, you will be able to reach out for and grasp any object within your reach that has sufficient visual contrast, be it a coffee mug, a door handle, a coin, and of course your white cane.

|  |
| --- |
|  |

Video 1 Mastering camera-hand coordination with The vOICe

# Interpreting distance and size

Images do not explicitly represent distance and size of objects in the view, but a number of indirect visual clues for distance and size exist that you can learn to exploit.

## Changes in apparent size with distance

The most powerful and general distance and size clue is how an object changes in apparent size as you move back and forth. Here too a very simple rule of visual perspective applies: *an object appears twice as large at half the distance*. This means that if an object was already close and already filled a substantial part of your view, it will fill a much bigger part of your view when moving only a little closer, because it does not take much to halve the distance when the object is already close. Also, when the object fills a large part of your view while its distance is small, the physical object size will be comparable to its physical distance (give or take a factor two, for a typical camera viewing angle of some 50 to 120 degrees). An object with a physical width or height of half a meter will typically span your view at about arm’s length. However, a much bigger object will span your view at a much larger distance, so apparent size in your view is not enough to know tell distances.

If you sit at the dark tabletop and put a bright coffee mug on it within arms reach, it will immediately appear noticeably bigger when you lean forward or move the mug towards you. On the other hand, a distant object such as a distant building, no matter how big it appears in your view, will barely change in apparent size when you make a few strides towards it. Because of its larger distance, it will now take many strides to halve the distance and make it appear twice as large in your view. Therefore, by judging the *amount of change* in apparent size as you move back and forth, you can judge the physical distance to the object. Keep in mind – and you must be thoroughly aware of this – that *apparent size as such says nothing about distance*: a building at say fifty meters distance can have the same rectangular appearance and same apparent visual size as a doorway at five meters distance, in case the building is physically ten times as tall and wide as that doorway. So you can never reliably judge distance from a single static view (unless you also *recognize* the object and know its physical size to relate apparent size to physical size, but that is much harder to master than the method that we describe here). You really need to move in order to reliably judge distances. This is an aspect of what is called “active vision”, because by moving around you get more information than by standing still, and that applies in particular to information about distance and size.

This presents a powerful method for use in both obstacle detection and navigation! If a static object becomes noticeably bigger after a few strides it must be nearby and therefore a potential obstacle that you may wish to avoid or at least anticipate before touching upon it with your cane. On the other hand, if a static object barely changes in apparent size after a few strides it is distant and therefore a potential landmark that you can make use of in navigation, for instance to maintain a desired heading, as discussed in the section on “Visual landmarks”.

What was said above also applies to equally-spaced vertical stripes. When the associated rhythm becomes twice as slow when making a single stride forward, the stripes must be quite close, such as with a bookshelf within arm’s reach, while if there is only a modest change in rate after a few strides there may be a fence at say ten meters distance.

Now practice this distance clue extensively with objects of various sizes and at various distances. The effects are often fairly subtle, and at least initially it takes a lot of conscious effort not to “overlook” some growing visual pattern embedded within the often complex cacophony of sounds in typical environmental soundscapes. Subtle clues easily drown out among less subtle signals until you learn to “listen out” the details that matter to you. When you get better at perceiving apparent changes this will be highly beneficial for your ability to move around, avoiding obstacles while tracking various landmarks that are more distant. Understanding the principles does not suffice. You must practice this on a daily basis, for at least ten minutes a day.

*Initially practice the perception of changes in apparent size when moving back and forth for half an hour daily, for at least two weeks.* Together with the reaching and grasping exercise, this will mean spending about an hour on daily training, for at least two weeks. After that, you can halve the duration of the exercises to fifteen minutes daily on reaching and grasping and another fifteen minutes on interpreting distance and size as you move back and forth among various objects at various distances. So you are from then on at half an hour daily practice, which is considered the minimum for making good progress, just like you’d need to practice at least half an hour daily to make good progress in mastering a musical instrument or a foreign language. You should sustain this level of daily training for at least a year. Of course it is a matter of trade-offs: to become a concert pianist requires hours of daily practice with a musical instrument, but it is just not realistic to expect this from everyone because of various other social or job obligations. You can certainly attain a decent level with half an hour daily practice, if you do this with good focus and a continued active interest in improving your vision skills.

## Other distance and size clues: parallax and occlusion

In addition to distance indications that you get from moving back and forth, you can obtain distance indications from moving sideways, through a phenomenon called *visual parallax*. Another visual effect is that anything that is behind an object from your viewpoint appears hidden through what *visual occlusion*. Still more subtle and implicit distance clues are provided by shading and shadows, although the latter will typically not offer very effective distance clues for use with The vOICe. Parallax and occlusion can be very useful though in supplementing the distance clues discussed in the section “[Changes in apparent size with distance](#_Ref366398020)”.

*Visual parallax*.

When moving sideways, objects in your view appear to move in the opposite direction, but the amount of apparent displacement depends on their distance. A distant background does not appear to move at all, and the apparent displacement of nearby (foreground) objects against a distant background tells you that those objects are indeed closer than the background. A ranking by amount of displacement gives you the distance order of objects.

For initial practice, go stand in front of a pole, with some visual pattern in the more distant background (such as buildings) and notice how soundscapes change as you do one step sideways and back. If you do this properly, that is, without turning at the same time, you will notice that the background appearance remains the same while the pole moves in the opposite direction of your step. Moreover, when the pole is farther away it will appear to move less with your single sideways step. If the background seems to move too then you are inadvertently turning around a bit, and vice versa, by keeping the background constant in your view you ensure that your heading does not change.

*Visual occlusion*.

Unless an object happens to be visually transparent, like with glass, you cannot see what is right behind it: what lies behind the object is occluded. This makes that more distant objects and parts of the background can appear hidden, in whole or in part. In combination with visual parallax this makes that what is hidden changes with sideways movements.

*Shading.*

The angle that a surface makes with the direction of a light source affects the apparent brightness from reflected light. An object such as a sphere that is curved towards you will therefore show a specific variation of brightness across its surface that depends on its three-dimensional shape, providing a distance (and shape) clue. It is hard to give explicit rules for this: it is something that you learn from extensive experience with various types of objects and their visual appearance under various lighting conditions.

*Shadows.*

Light originating from a directional light source may be occluded by objects, causing surface areas behind the object to appear darker. These darker areas are called shadows, cast by objects. Like with shading, the shadows of objects give implicit information about the three-dimensional placement of objects in the environment, and thus provide supplemental distance (and shape) clues.

# Visual perspective

One effect of visual perspective you have already learnt in the context of interpreting distance and size, namely that an object appears twice as large at half the distance. However, this effect has a range of related consequences. For example, when you look along a road into the distance, the road appears twice as narrow at twice the distance, and at very large distances the apparent width of the road becomes vanishingly small. This is why people sometimes speak of so-called vanishing points. It makes a long straight road appear as an upright triangle, where the top corner corresponds to the vanishing point. If there are rows of buildings lining both sides the road, these give a rhythm in the left-to-right soundscape scan that first speeds up as the row of buildings is traced into the distance towards the vanishing point, and then slows down as the row of buildings on the right side of the road is traced until the soundscape ends with nearby buildings. So the rhythm becomes faster at larger distances and slower at smaller distances, and since buildings are often roughly comparable in physical size and shape this too can give you a clue about distance if you already know that you are on a road. Now this is not easily applied as a training exercise, so we will instead look at similar effects of visual perspective when looking at a computer keyboard. Find a regular computer keyboard for use in the next exercise, and look at it from the top side.

The rows of keys on the keyboard give a characteristic rhythm. This rhythm slows down if you move the camera closer to the keyboard, and it speeds up as the distance gets larger, just as with the earlier bookshelf example. This is because visual items appear bigger at close range, such that fewer keys fit in the view. Now look at the keyboard under an angle. You will notice that the rhythm speeds up or slows down within one visual sound view, because the more distant parts of the keyboard have a faster rhythm than the closer parts. How exactly the rhythm speeds up or slows down depends on the orientation of keyboard with respect to the camera. Apply various angles and distances until you are completely familiar with how the soundscapes change with distance and orientation.

[](file:///T:\HTML\vOICe\manual\keyboard_topview.mp3) [](file:///T:\HTML\vOICe\manual\keyboard_topcloseup.mp3) [Computer keyboard seen from above and on the left side
](file:///T:\HTML\vOICe\manual\keyboard_leftslanted.mp3) [](file:///T:\HTML\vOICe\manual\keyboard_rightslanted.mp3)

Note that it also works the other way around: from hearing how fast the keyboard rhythm is and how it speeds up or slows down you can tell both the distance and orientation. The same effects apply to rows of buildings along a road and rows of windows in each building, only here it may take a number of strides before you notice changes in the soundscape, because of the larger distances involved.

# Visual landmarks

Characteristic visual patterns in the distant background can be used as visual landmarks. Because of their distance, landmarks change only slowly in visual appearance as you walk ahead. They can help to maintain a constant heading. For example, if you are in an open parking lot, keeping the pattern of a distant building at a constant position in the soundscapes can work much like a visual “compass”. If the visual patterns are sufficiently unique they can also help you to know where you are along a route without a need for counting steps. Some building facades may have pillars that give a characteristic rhythm; others may have a particular arrangement of windows that make them stand out among the rest, while some shops may have the shop name in large letters above the entrance. After all, most companies want to stand out and be noticed. Try to remember what things give characteristic soundscapes in your environment, such that over time you will at all positions along a familiar route know where you are just by looking around and noticing patterns that are more or less unique to that position along the route. Of course this will not work for city blocks that all look alike.

# Ground level hazards

|  |
| --- |
| Schematic illustration (for sighted trainers) showing how to adjust your viewing angle as you move closer to a ground level obstacle. |

Although nothing beats the white cane in reliably detecting nearby ground level hazards such as obstacles, step downs or pot holes, proper visual strategies can certainly help too, especially in anticipating potential hazards well before hitting upon them with the cane, even letting you walk around those hazards without ever touching upon them with the cane by turning left or right in time. The latter possibility is not further discussed here. However, we will consider in more detail the situation where you keep walking towards a ground level obstacle until it becomes a tripping hazard unless detected at the last second by the cane.

In general, while walking it is best to look slightly downward. The optimal viewing angle is usually not straight ahead (unless you are dealing with distant landmarks), because items above head level rarely present collision threats. Instead, the horizon, corresponding to head level, should normally show near the top of your camera view instead of near the middle. This still lets you detect head level hazards while letting you see more of the ground in front of you, at closer range than when looking straight ahead. The best viewing angle depends on the field of view of your camera, but a 20 degree downward tilt may be assumed as a ballpark figure for a camera with a 45 degree vertical field of view. Then, once you detect a potential ground level hazard, you very gradually increase your downward head tilt in order not to let the hazard leave your view at the bottom side, rendering it invisible, and you track it by keeping it near the bottom of your view as you move closer. The changes in apparent visual size in combination with your increasing head tilt will tell you how close you are and will let you avoid tripping over it in case the hazard is a ground level object. It is advisable to occasionally briefly look up to make sure that in the mean-time no head level hazards showed up, because you would otherwise not notice those due to the increasing downward tilt while tracking the ground level hazard.

# Training schedule

As outlined in the preceding sections, your recommended minimum training schedule is as follows.

Weeks 1 and 2:

* + 30 minutes of daily training for “[Reaching and grasping](#_Ref366322639)”.
  + 30 minutes of daily training for “[Interpreting distance and size](#_Ref366322958)”.

Weeks 3 and later, for at least one year:

* + 15 minutes of daily training for “[Reaching and grasping](#_Ref366322639)”.
  + 15 minutes of daily training for “[Interpreting distance and size](#_Ref366322958)”.
  + Use The vOICe in daily living to an extent that suits you.

More training time than proposed here can be helpful, but do not exaggerate and do not go above spending twice the recommended effort. However, there is no real upper limit to the time you can spend using The vOICe in your normal daily living activities, just like the sighted use their eyes all day. Social or job obligations will often be limiting factors that one will have to strike a balance with.

# Performance checklist

Especially in self-training, it is vital that you are very critical about your own performance and make sure that you reach the intended performance levels to lay a solid basis for further use of The vOICe. So how do you know that you have reached the goals of this training manual? For that you must convince yourself that you can state the following to approach sighted performance:

* After simultaneously and randomly casting two bright objects – such as DUPLO bricks – on a dark tabletop, I can grab both of them directly and within three soundscapes, without sweeping of my hands to correct a near miss. I get this right at least half of the time. This proves that I have mastered good camera-hand coordination.
* In a familiar (home) environment that I used to practice mobile use of The vOICe, I can walk now around freely without touching any walls, door posts or furniture. I can take steps forward and backward to judge my distance to nearby objects from the changes in their apparent size. In doing so I can next each out and touch the object at a specific point such as an edge without trial and error.
* Starting at one end of a familiar room, I can walk toward a salient item at the other end of the room, stopping once fast changes in apparent size in the soundscapes indicate that the object is within arm’s reach. Then I can reach out and touch the object at a specific point. The salient object can for instance be a painting on the far wall, or a doorpost of a doorway at the other end of the room.
* Starting in the middle of a familiar room I can turn myself around several times and easily reorient myself from the soundscapes and walk in a specific direction, for instance towards a doorway.
* When I drop my keychain on a dark carpet, I can easily locate it using The vOICe, bend over and fetch it in one smooth directed movement, without any groping.

Always remember that if you are still performing a basic visual task slowly after training, you are not doing it right yet: you should be able to do it about as fast as the sighted. *Keep pushing yourself for speed, or you may become stuck at a slow conscious analysis level and not reach your full potential (nor that of The vOICe).*